

Introduction to JuiceFS

[JuiceFS](#) is an open-source, high-performance distributed file system designed for cloud environments. Released under the Apache License 2.0, it offers full POSIX compatibility, enabling various object storage services (such as Amazon S3) to function as massive, shared local disks. JuiceFS allows seamless mounting and access across different hosts, platforms, and regions.

Architecture

JuiceFS separates data and metadata storage:

- Data: Files are split into chunks and stored in object storage.
- Metadata: Managed in databases such as Redis, MySQL, TiKV, and SQLite, depending on performance and availability needs.

This design ensures scalability, high availability, and low maintenance.

Key features

- POSIX compatibility – Works like a local file system, easy to integrate with existing applications.
- HDFS compatibility – Supports the HDFS API for improved metadata performance.
- S3 Gateway – Provides an S3-compatible API for object storage access.
- Cloud-native – Easy to use in Kubernetes via the CSI driver.
- Distributed architecture – Can be mounted on thousands of servers with concurrent read/write access.
- Strong consistency – File changes are immediately visible across all **SERVERS**.
- High performance – Achieves millisecond-level latency and scales throughput with the object storage backend.
- Data security – Supports encryption in transit and at rest.
- File lock – Supports BSD (flock) and POSIX (fcntl) locks.
- Data compression – Supports LZ4 and Zstandard to optimize storage usage.

Scenarios

JuiceFS is designed for large-scale data storage and serves as a powerful alternative to alternative to many distributed file systems and network file systems, especially in the following scenarios:

Big data analytics

- Fully compatible with HDFS.

- Seamlessly integrates with Spark, Presto, Hive, and other data processing frameworks.
- Offers better performance compared to direct object storage access.

Machine learning

- POSIX-compliant, supporting all ML/DL frameworks.
- Enables shared file storage for collaborative model training and dataset management.

Kubernetes

- CSI support allows decoupled persistent storage for stateless applications.
- Ideal for multi-container data sharing.

Shared workspaces

- Supports multi-host concurrent read/write operations.
- ensures smooth data flow and supports scripting operations.

Data backup

- Provides scalable storage space for backing up all kinds of data.
- With its shared mount feature, data from multiple hosts can be aggregated into one place and then backed up together.